

Actividad 4: Aplicar el modelo de correlación lineal en censos de la población

PROPÓSITO

Los estudiantes profundizan en la correlación de dos variables y el modelo de regresión lineal. Se propone que primero usen herramientas tecnológicas para abordar una cantidad acotada de datos (regiones de Chile), para extender el trabajo posteriormente a una base de datos más amplia (comunas del país).

Objetivos de Aprendizaje

OA 2. Resolver problemas que involucren los conceptos de media muestral, desviación estándar, varianza, coeficiente de variación y correlación muestral entre dos variables, tanto de forma manuscrita como haciendo uso de herramientas tecnológicas digitales.

OA b. Resolver problemas que impliquen variar algunos parámetros en el modelo utilizado y observar cómo eso influye en los resultados obtenidos.

OA d. Argumentar, utilizando lenguaje simbólico y diferentes representaciones para justificar la veracidad o falsedad de una conjetura, y evaluar el alcance y los límites de los argumentos utilizados.

OA i. Buscar, seleccionar, manejar y producir información matemática/cuantitativa confiable a través de la web.

Actitudes

- Trabajar colaborativamente en la generación, desarrollo y gestión de proyectos y la resolución de problemas, integrando las diferentes ideas y puntos de vista.
- Interesarse por las posibilidades que ofrece la tecnología para el desarrollo intelectual, personal y social del individuo.

Duración: 12 horas pedagógicas

DESARROLLO

EL VALOR DE LA RELACIÓN ENTRE DENSIDAD Y CANTIDAD DE HOGARES EN LAS REGIONES DE CHILE

El censo es el conteo y la caracterización de todas las viviendas y habitantes del territorio nacional en un momento determinado. Hay censos de población y censos de vivienda. Los primeros permiten contabilizar y caracterizar a la población en un momento del tiempo: magnitud, distribución y composición de la población (sexo, edad, fertilidad, educación, migración); con los segundos se puede contabilizar e indagar en las características de las viviendas y hogares donde residen los habitantes.

Conexión interdisciplinaria:
Educación Ciudadana
OA a, c, 3° y 4° medio

1. Descarga la hoja de cálculo “Censo2017_Densidad_Hogares.xlsx” con los resultados de la “densidad de la población” y la “cantidad de hogares” de las 16 regiones y 346 comunas de Chile, desde:

https://www.curriculumnacional.cl/link/https://www.dropbox.com/s/7334chx8pxmkej/Censo2017_Densidad_Hogares.xlsx?dl=0

Región	Densidad de la población (km ²)	Cantidad de hogares
Región de Tarapacá	7,82	97693
Región de Antofagasta	4,82	174314
Región de Atacama	3,78	88706
Región de Coquimbo	18,67	240317
Región de Valparaíso	111,27	608949
Región del Libertador Gral. Bernardo O’Higgins	55,93	301717
Región del Maule	34,47	351509
Región de Biobío	64,95	507241
Región de la Araucanía	30,08	317525
Región de Los Lagos	17,11	277482
Región de Aysén del Gral. Carlos Ibáñez del Campo	0,96	34609
Región de Magallanes y de la Antártica Chilena	0,12	53832
Región Metropolitana de Santiago	462	2238179
Región de Los Ríos	21,07	128984
Región de Arica y Parinacota	13,4	66842
Región del Ñuble	36,67	163738
País	8,77	5651637

2. Considera las siguientes definiciones:

Densidad de población es la relación entre el número de personas que viven en un territorio y su extensión, lo que normalmente se expresa en habitantes por km^2 .

Habitantes, para efectos del censo, se considera como a las personas efectivamente censadas en el territorio.

Hogar como una o más personas que, unidas o no por un vínculo de parentesco, alojaron la noche del 18 al 19 de abril de 2017 en una misma vivienda o parte de ella y se benefician de un mismo presupuesto para alimentación.

- a. Con la “Hoja de cálculo” de GeoGebra, construyan el gráfico de nube de puntos (herramienta “Análisis de Regresión de dos variables”), considerando como variable independiente la densidad de población, como variable dependiente la cantidad de hogares, y a las 16 regiones de la pestaña de Excel llamada “Región” (recuerden excluir los valores totales para el país al final).

- b. Basados en el gráfico anterior, agreguen la recta de regresión a dicho gráfico (“Modelo de regresión Lineal”) y describan el tipo de relación lineal existente entre las variables.
3. Completen la siguiente tabla, usando los valores de la planilla (pestaña Región).

Región	Densidad de P. (X)	C. de Hogares (Y)	$X \cdot Y$	X^2	Y^2
I	7,82	97693	763959,26	61,1524	9543922249
II	4,82	174314			
III	3,78	88706			
IV	18,67	240317			
V	111,27	608949			
VI	55,93	301717			
VII	34,47	351509			
VIII	64,95	507241			
IX	30,08	317525			
X	17,11	277482			
XI	0,96	34609			
XII	0,12	53832			
XIII	462	2238179			
XIV	21,07	128984			
XV	13,4	66842			
XVI	36,67	163738			

- a. Según la relación entre las variables, ¿cómo debería ser el valor del coeficiente de correlación? Compartan ideas en el grupo y elaboren una hipótesis o conjetura.
- b. Calculen la suma de todos los valores de las columnas ($X, Y, X \cdot Y, X^2, Y^2$) y obtengan el coeficiente de correlación lineal de Pearson.
- c. Utilizando la calculadora científica o la función de Excel “COEF.DE.CORREL()”, comprueben el coeficiente de correlación obtenido. Si usan la calculadora científica, recuerden escoger modo REG LIN, ingresar los valores separados por comas y guardar los datos con la tecla M+.
4. Comparen el tipo de relación lineal mostrado en el gráfico y el valor del coeficiente de correlación obtenido.
- a. ¿Se confirma su conjetura? Argumenten.
- b. ¿Qué pueden decir del valor de correlación obtenido? Compartan ideas en el grupo.
5. Analicen en el grupo las siguientes situaciones.
- a. Supongan que se crea una nueva región para el país, cuya densidad es de $9,53 \frac{\text{hab}}{\text{km}^2}$. ¿Cuál sería la cantidad de hogares estimada para dicha región, bajo el modelo encontrado?
- b. Si en la XIII Región aumentara la densidad de la población en $35,4 \frac{\text{hab}}{\text{km}^2}$, ¿en cuánto debería aumentar la cantidad de hogares según el modelo de regresión lineal?
- c. ¿En qué regiones del país sería factible hacer una subdivisión de la población, de acuerdo con la densidad de $9,53 \frac{\text{hab}}{\text{km}^2}$? Fundamenta en grupo esta respuesta.

EL VALOR DE LA RELACIÓN ENTRE DENSIDAD Y CANTIDAD DE HOGARES EN LAS COMUNAS DE CHILE

- Con base en el contexto y las definiciones de variables entregadas en la actividad anterior, y usando una planilla Excel, respondan lo siguiente:
 - Construyan –con la “Hoja de Cálculo” de GeoGebra– el gráfico de nube de puntos (herramienta “Análisis de Regresión de dos variables”), considerando como variable independiente la densidad de población y como variable dependiente, la cantidad de hogares de las 346 comunas de la pestaña de Excel “Comunas”.
 - Agreguen la recta de regresión (“Modelo de regresión Lineal”) al gráfico y describan el tipo de relación lineal existente entre las variables.
 - Según la relación entre las variables, ¿cómo debería ser el valor del coeficiente de correlación? Compartan ideas en el grupo y elaboren una hipótesis o conjetura.
 - Utilizando la calculadora científica o la función de Excel “COEF.DE.CORREL()”, calculen el coeficiente de correlación.
- Comparen el tipo de relación lineal mostrado en el gráfico y el valor del coeficiente de correlación obtenido.
 - ¿Se confirma su conjetura? Argumenten.
 - ¿Qué pueden decir del valor de correlación obtenido? Compartan ideas y argumenten.
- Se necesita saber la correlación entre las comunas de cada una de las regiones del país.
 - Completen las siguientes tablas con la información que se solicita:

	Densidad de población (X)	Cantidad de hogares (Y)
Comuna a la que pertenece tu establecimiento educacional		

Usa el Modelo de:	Modelo de regresión lineal ($\hat{Y} = \beta_0 X + \beta_1$)	Valor de coeficiente de correlación (r)	Cantidad de hogares estimada de la comuna a que pertenece el establecimiento (\hat{Y})	Diferencia entre valor real y valor estimado de cantidad de hogares ($Y - \hat{Y}$)
16 regiones				
346 comunas				
Comunas pertenecientes a la región de tu establecimiento educacional				

- b. ¿Qué región tiene correlación más cercana a 1 entre sus comunas? Compartan ideas y argumenten.
 - c. ¿Qué región tiene correlación más cercana a 0 entre sus comunas? Compartan ideas y argumenten.
1. Analicen en el grupo las siguientes situaciones y justifiquen en cada caso.
 - a. Que el coeficiente de correlación sea más cercano a 1, ¿implica necesariamente que la predicción sea mejor?
 - b. ¿A qué se debe que la estimación sea mejor en unos casos que en otros?
 - c. ¿Qué modelo de regresión lineal se ajusta más a la realidad de la comuna de tu establecimiento?
 - d. Si la densidad de población de la comuna en que se encuentra su colegio aumentara al doble, ¿en cuánto variaría la cantidad de hogares con cada modelo calculado en la tabla realizada en 3.a?
 - e. ¿Qué modelo aplicarían para predecir el aumento de la cantidad de hogares de la pregunta anterior?

EXPLORANDO MÁS LA RELACIÓN ENTRE DENSIDAD Y CANTIDAD DE HOGARES

1. Según el contexto y las definiciones de variables entregadas en la actividad anterior, y usando planilla Excel, respondan lo siguiente:
 - a. Si calculamos el coeficiente de correlación entre la región a la que pertenece su colegio y las regiones que colindan por el sur y el norte, ¿mejora el valor del coeficiente respecto del coeficiente de correlación de solo la región del establecimiento? Si está situado en las regiones extremas, usen los valores de las dos regiones siguientes hacia el norte o hacia el sur.
 - b. Grafiquen la nube de puntos de los datos del ejercicio anterior.
 - c. ¿Qué otro tipo de regresión que no sea lineal se podría utilizar? Apoyen su decisión, usando GeoGebra.
2. Analicen las siguientes situaciones y justifiquen en cada caso.
 - a. Si la densidad poblacional disminuyera a la mitad, ¿cuál sería la cantidad de hogares estimada con el modelo de regresión lineal de los datos anteriores?
 - b. Si la densidad poblacional disminuyera a la mitad, ¿cuál sería la cantidad de hogares estimada con el nuevo modelo de regresión lineal? Comparen lo obtenido con los datos anteriores y con el modelo de regresión lineal.
3. Según sus observaciones, ¿cuál sería el mejor modelo de regresión?

ORIENTACIONES PARA EL DOCENTE

1. Si se cuenta con los recursos tecnológicos, se sugiere que los estudiantes interactúen entre GeoGebra y la planilla de cálculo. Para traspasar los datos, sólo se necesita copiar los datos desde Excel y pegarlos en la hoja de cálculo de GeoGebra.
2. Conviene trabajar con los alumnos paso a paso en GeoGebra. Por ejemplo, para crear el gráfico, se selecciona las columnas de la variable independiente y dependiente a graficar, luego se pincha (clic) en la herramienta de GeoGebra “Análisis de Regresión de dos variables” (ubicada en el menú de cuadrados de la hoja de cálculo). Se despliega un recuadro donde aparecerán los datos a graficar.
3. Si se copió el encabezado dentro de la columna, se recomienda marcar el ícono de tuerca para seleccionar la opción “Encabezado(s) como título(s)” y pinchar “Analiza”: aparecerá un recuadro con el gráfico de puntos. Para agregar la recta de regresión, sólo hay que hacer clic debajo del título “Modelo de regresión” y seleccionar la opción “Lineal”.
4. Para exportar a la vista gráfica de GeoGebra, hay que pinchar en la vista de la nube de puntos con el botón derecho del mouse, y seleccionar “copiar en Vista Gráfica”. Si se desea exportar el resultado de la regresión como imagen, seleccionar “Copiar al portapapeles” (guardará la imagen directamente y se podrá pegar en el procesador de texto) o “Exportar como imagen” (creará un archivo de imagen).

5. Dada la naturaleza de las actividades, conviene que los jóvenes también obtengan los valores con calculadora. Habría que trabajar paso a paso, ya que es la oportunidad para mostrar de cómo se obtiene el valor del coeficiente de correlación.
6. Cuando se necesite calcular cada uno de los coeficientes de correlación de las comunas por regiones, se sugiere utilizar la fórmula de Excel “COEF.DE.CORREL()” y luego comparar cada uno de sus resultados para responder a las preguntas.
7. Para decidir cuál modelo es mejor, pueden comparar los valores que se obtiene al reemplazar el valor de la densidad de la población en cada modelo de regresión, y luego compararlos con el valor real de su contexto. Dado que se calculó en la tabla, el que esté más cercano al valor real será el más aproximado, considerando un posible error que oscila entre las diferencias obtenidas en las comparaciones.
8. Cabe enfatizar que lo que mide el valor del coeficiente de correlación es que haya una correlación fuerte o débil entre las variables, pero no indica que el modelo empleado sea necesariamente el mejor. Además, es importante el contexto en que se aplique el modelo.
9. Se sugiere los siguientes indicadores para evaluar formativamente los aprendizajes:
 - Toman decisiones fundamentadas respecto de la correlación entre variables, dependiendo del coeficiente o factor de correlación.
 - Resuelven problemas que involucran la correlación muestral entre dos variables.

RECURSOS Y SITIOS WEB

Sitios web sugeridos para estudiantes y profesores

- “Censo2017_Densidad_Hogares.xlsx”
https://www.curriculumnacional.cl/link/https://www.dropbox.com/s/7334chx8pxmkej/Censo2017_Densidad_Hogares.xlsx?dl=0
- Resultados Censo 2017
<https://www.curriculumnacional.cl/link/http://resultados.censo2017.cl/>
- ¿Qué es y cómo se interpreta el coeficiente de correlación de Pearson?
<https://www.curriculumnacional.cl/link/https://www.webyempresas.com/coeficiente-de-correlacion-de-pearson/>
- Regresión lineal
<https://www.curriculumnacional.cl/link/https://www.GeoGebra.org/m/ppCJqUuf>
- Video: Cálculo del coeficiente de correlación lineal
<https://www.curriculumnacional.cl/link/https://www.youtube.com/watch?v=7ifeqKgZyHA>